Helmet detection using Python with AI

P.Muthayulu , SK.Arshad
Asst.Proffessor ,Student
Narayana Engineering College,Gudur,India

Abstract: In traffic accidents, motorcycle accidents are the main cause of casualties, especially in developing countries. The main cause of fatal injuries in motorcycle accidents is that motorcycle riders or passengers do not wear helmets. In this paper, an automatic helmet detection of motorcyclists method based on deep learning is presented. The method consists of two steps. The first step uses the improved YOLOv5 detector to detect motorcycles (including motorcyclists) from video surveillance. The second step takes the motorcycles detected in the previous step as input and continues to use the improved YOLOv5 detector to detect whether the motorcyclists wear helmets. The improvement of the YOLOv5 detector includes the fusion of triplet attention and the use of soft-NMS instead of NMS. A new motorcycle helmet dataset (HFUT-MH) is being proposed, which is larger and more comprehensive than the existing dataset derived from multiple traffic monitoring in Chinese cities. Finally, the proposed method is verified by experiments and compared with other state-of-the-art methods. Our method achieves mAP of 97.7%, F1-score of 92.7% and frames per second (FPS) of 63, which outperforms other state-of-the-art detection methods.

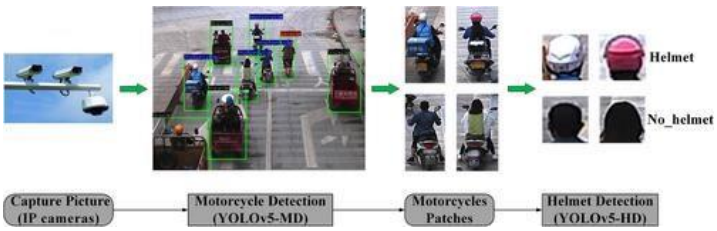
Keywords: Automatic helmet detection*;* Deep learning*;* YOLOv5 detector*;* Triplet attention*;*

INTRODUCTION

According to the latest report of 'Global Road Safety Status in 2018' released by the World Health Organization (WHO) [1], about 1.35 million people die in road traffic accidents every year, among which 28% die from motorcyclists. Especially in some underdeveloped areas, due to the restrictions of urban infrastructure and economic conditions, motorcycles have become the main tool of transportation, and the death rate of road traffic in these areas is about three times that in developed areas. In Southeast Asia and the Western Pacific region, such as India, Vietnam, Indonesia and other countries, motorcycle traffic accident deaths accounted for 43% and 36% of all traffic accident deaths respectively. The WHO points out that the head injury of motorcyclists is the main cause of death. If motorcyclists wear helmets correctly, the risk of death can be reduced by 42%, and the risk of head injury can be reduced by 69%. Therefore, motorcyclists must wear helmets [1]. However, in some developing countries, the rate of wearing helmets has been very low for various reasons. For example, according to the data of Thailand road foundation [2], motorcycle traffic accidents in Thailand cause about 5500 deaths every year, and only 20% of passengers in the back seat of motorcycles wear helmets. In order to increase the use of helmets, the Indian government has proposed various penalties under the Motor Vehicles (Amendment) Law of 2019. Under section 194d, a motorcyclist who does not wear a helmet will be fined 1000 rupees and disqualified for the driving license for three months. Even if these laws exist, people will still try to escape the arrest of traffic police, and strict law enforcement also requires a lot of police, which is time-consuming and costly. Therefore, it is necessary to develop an automatic helmet detection of motorcyclists system based on deep learning to reduce the

number of deaths in motorcycle traffic accidents. In recent years, with the rapid development of deep learning, convolutional neural network (CNN) has been widely used, such as semantic segmentation, object detection, all of which have made great breakthroughs. Semantic segmentation, a pixel-level vision task, is developed rapidly by using CNNs. Wang et al. [3] propose a weakly supervised adversarial domain adaptation to improve the segmentation performance from synthetic data to real scenes. At present, object detection methods based on deep learning are generally divided into two categories: one-stage algorithms and two-stage algorithms. The two-stage algorithms are represented by the faster R-CNN [4] and the mask R-CNN [5]. These algorithms have high detection accuracy, but the speed is very slow, which makes it difficult to realize real-time detectionFUNCTIONAL OVERVIEW

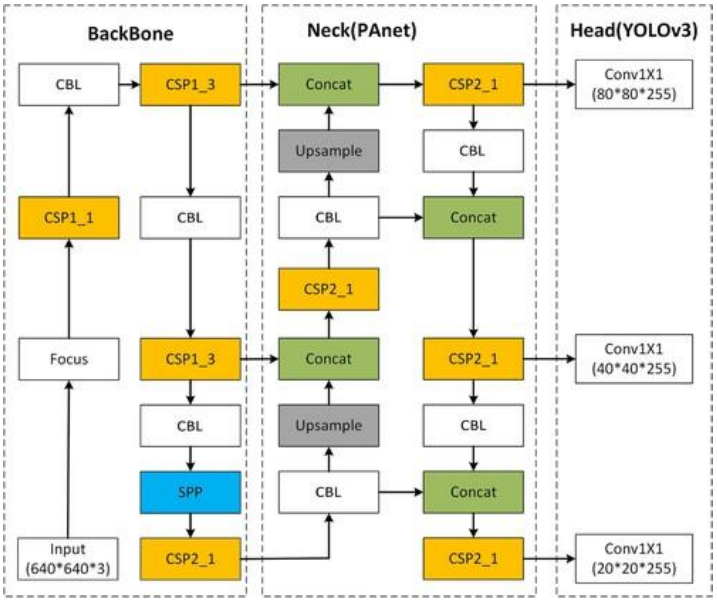
On start after giving In urban traffic, there are many kinds of vehicles on the road, such as two-wheeled vehicles, three-wheeled vehicles, four-wheeled vehicles, and road congestion often occurs. In such a complex scene, it is a very challenging task to accurately detect the motorcycle and judge whether the rider on the motorcycle is wearing a helmet. Although a variety of helmet detection of motorcyclists methods have been proposed in some literature, these methods have many shortcomings, such as the limitations of accuracy and speed of traditional methods, and lack of high-quality traffic monitoring scene datasets. In this section, we propose a real-time and accurate automatic helmet detection of motorcyclists method based on deep learning, which includes two steps, as shown in Figure 2. The first step is motorcycle detection. First, the image to be detected is obtained from the video surveillance, and then the improved the YOLOv5 algorithm, namely YOLOv5-MD, is used to detect the riding motorcycle in the image, which contains at least one rider. The second step is helmet detection, which takes the motorcycle region detected in the first step as the input of the second step, and then continues to use the improved YOLOv5 algorithm, namely YOLOv5-HD, to detect whether the motorcyclists are not wearing helmets. Because the tasks of motorcycle detection and helmet detection are quite different, the network is designed for each stage, that is, YOLOv5-MD and YOLOv5-HD, in order to better improve the detection performance. Section 5 introduces the proposed network, Section 3.2 introduces the motorcycle detection method, and Section 3.3 introduces the helmet detection method.



3.1 The proposed network

3.1.1 YOLOv5 network

YOLO is a classical one-stage object detection algorithm. It turns the detection problem into a regression problem. Instead of extracting RoI, it directly generates the bounding box coordinates and probability of each class by the regression method. Compared with faster R-CNN, it greatly improves the detection speed. In 2020, the fifth version of YOLO was proposed by utralytics and named YOLOv5, which surpasses all previous versions in speed and accuracy. The YOLOv5 algorithm uses the parameters depth_multiple and width_multiple to adjust the width and depth of the backbone network, so as to get four versions of the model, which are YOLOv5s, YOLOv5m, YOLOv5l, YOLOv5x. YOLOv5s is the simplest version with the smallest model parameters and the fastest detection speed. Its network structure is shown in Figure 3, which is composed of focus, Conv-Bn-Leakyrelu (CBL) and CSP1_x, CSP2_x and spatial pyramid pooling (SPP) modules. The focus block mainly contains four parallel slice layers to tackle with the input image. The CBL block contains a convolutional layer, batch normalization layer and hard-wish function. CSP1_x block contains CBL blocks and x residual connection units. CSP2_x block contains x CBL blocks. SPP block mainly contains three maxpool layers.



YOLOv5s divides the model into three parts: the backbone network part, the feature enhancement part and the head part. Each part has different functions.

The backbone network part is used to extract image features. First, the focus structure is used to extract pixels from high-resolution images periodically and reconstruct them into low-resolution images. That is to say, the four adjacent positions of the images are stacked, and the information of WH dimension is focused into the C channel space to improve the receptive field of each point and reduce the loss of original information. The design of the module is mainly to reduce the amount of calculation and speed up. Then, drawing on the design idea of CSPNet [43], the CSP1_x and CSP2_x modules are designed. The module first divides the feature mapping of the basic layer into two parts, and then combines them through the cross-stage hierarchical structure, which reduces the amount of calculation and ensures accuracy. In the last part of the backbone network, using the SPP network, this module can further expand the receptive field and help to separate contextual features.

The feature enhancement part is used to further improve the feature extraction ability. It uses the idea of PANet [44] to design the structure of FPN+PAN. First, it uses the FPN structure to convey strong semantic features from top to bottom, and then uses the feature pyramid structure constructed by the PAN module to convey strong positioning features from bottom to top. Through this method, it is used to fuse features between different layers.

The head part inherits the head structure of YOLOv3, which has three branches. The prediction information includes object coordinates, category and confidence. The main improvement is to use complete intersection over union (CIOU) loss as the bounding box region loss.

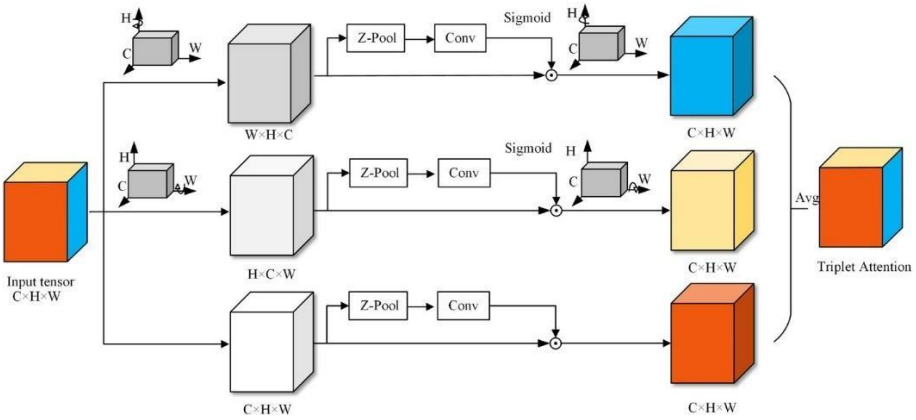
3.1.2 YOLOv5 with an attention mechanism

In motorcycle detection stage, due to the similar front appearance of motorcycles, tricycles and bicycles in riding state, the gap between the three different classes is relatively minor, while the differences between motorcycle samples are large because of the different postures, scales of motorcycles. Therefore, the large within-class gap and the large between-class gap lead to a great amount of false detection in the first stage. In the helmet detection stage, there are black helmets and black hair, hats and helmets, which are similar in colour and shape, often leading to false detection. In order to solve these difficult samples, we introduced the attention mechanism to optimize the feature extraction ability of the network, and effectively improve the detection accuracy.

In recent years, attention mechanisms have been applied to various tasks of computer vision. Non-local [45]

captures long-range features by calculating dependencies between features in different locations. SENet [46] proposed channel attention, which explicitly modelled dependencies between channels. GCNet [47] optimizes Non-local, combines global attention with channel attention. SKNet [48] uses a channel attention mechanism to achieve a dynamic selection of field perception. Convolutional block attention module (CBAM) [49] combines channel and spatial attention mechanisms in sequence mode.

We experimented with several different attention mechanisms and chose to add triplet attention [50] to the last layer of the backbone network. Attention modules that fuse both spatial and channel-wise attention usually model spatial and channel-wise dependencies separately which leads to a lack of semantic interaction between different dimensions of features. Triplet attention extracts the semantic dependence between different dimensions, eliminates the indirect correspondence between channels and weights, and achieves the effect of improving accuracy with little computational overhead. Figure 4 shows the basic structure of triplet attention. Triplet attention uses three parallel branching structures, two of which extract the inter-dimensional dependencies between two spatial dimensions and channel dimension C, and the other extracts the spatial feature dependencies. In the first two branches, triplet attention rotates the original input tensor 90° counter-clockwise along the H-axis and W-axis respectively, and transforms the shape of the tensor from C × H × W to W × H × C and H × C × W. In the third branch, the tensor is entered in its original shape C × H × W. After that, the tensor of the first dimension is reduced to the second dimension through the Z-pool layer, and the average aggregation feature and the maximum aggregation feature are connected.



Then, the reduced tensor is passed through the standard convolution layer with a kernel size of K, batch normalization layer, and finally, the attention weight of the corresponding dimension generated by the sigmoid function is added into the rotated tensor. At the final output, the output of the first branch rotates 90 degrees clockwise along the H axis and the output of the second branch rotates 90 degrees clockwise along the W axis, ensuring the same shape as the input. Finally, the output of the three branches is aggregated

equally as the output. The output tensor is defined as:

$$y = \frac{1}{3}\left(\overline{\widehat{x}_1\sigma\left(\varphi_1\left(\widehat{x_1^*}\right)\right)} + \overline{\widehat{x}_2\sigma\left(\varphi_2\left(\widehat{x_2^*}\right)\right)} + x\sigma\left(\varphi_3\left(\widehat{x_3^*}\right)\right)\right).$$

3.1.3 Soft-NMS for final processing of the results

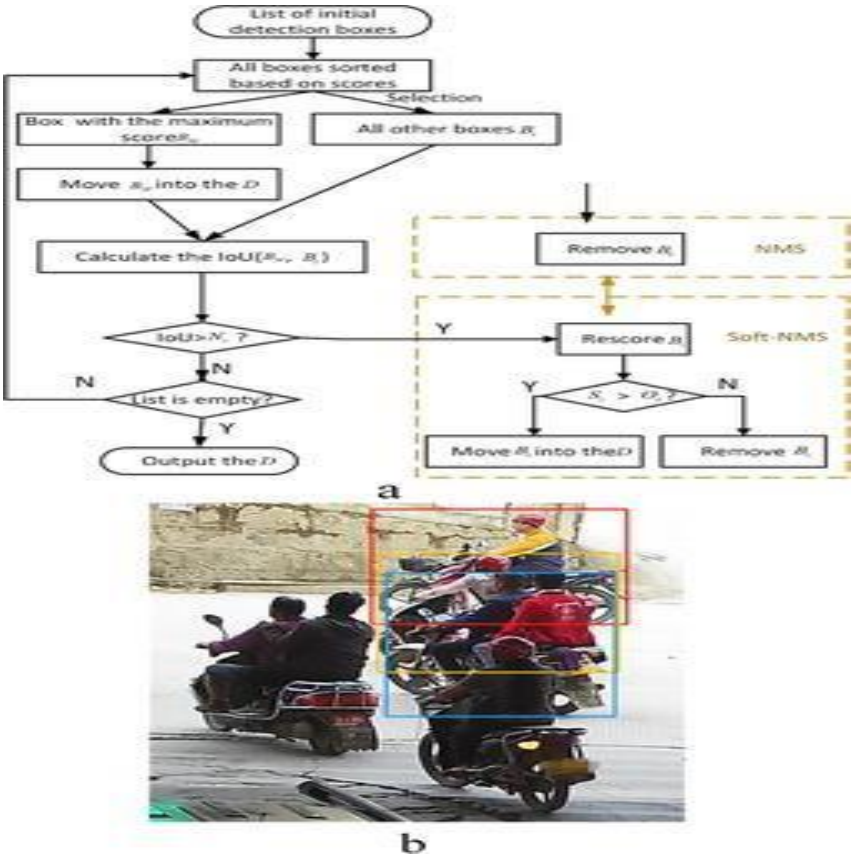
NMS is applied to most state-of-the-art detectors to obtain the final results because it significantly reduces the number of false positives. The flow chart of the NMS algorithm is shown in Figure 5a. First, all the detection boxes in the list are sorted according to their confidence scores. Second, the detection box

with the highest score is moved to the final detection list D, and the remaining detection boxes are assigned a unique identifier . Third, any prediction box whose overlap area with is greater than a certain threshold is removed. Repeat this process for the remaining boxes until the initial list is empty. However, motorcycles block each other on crowded roads, and the NMS algorithm will lead to missed detection. This problem is illustrated in Figure 5b. The blue and green boxes represent the detection results of different objects and correspond to different confidence scores. Suppose the blue box gets the highest score, followed by green and red. If the NMS algorithm is used, the green box will be removed because it has a large overlap with the blue box. Therefore, the motorcycle predicted by the green box will be ignored. So we introduced Soft-NMS instead of NMS to process the detection results. The NMS algorithm can be represented by rescoring function where is the score of the prediction boxis the predictionbox with the highest score, and is the overlap threshold. In NMS, a hard threshold is set to determine which boxes should be retained and which boxes should be deleted in domain. If an object actually exists but has an overlap rate with more than, its detection will be ignored. The core idea of soft-NMS is to use a penalty function to attenuate the scores of prediction boxes that overlap with rather than setting these scores to zero. The of soft-NMS is shown in the following formula

where

is a Gaussian penalty function, and σ is a super parameter selected according to experience. It is obvious thatthe score of the prediction box with a large overlap with will be greatly reduced, while the detection box far away from will not be affected. If the score of the prediction box is still higher than the confidence threshold

after penalty. Then the prediction box will be retained rather than discarded. Using soft-NMS to deal with motorcycles in crowded scenes can greatly reduce the missed detection rate of motorcycles.



MOTORCYCLE DETECTION

At present, YOLO series algorithms have been At present, YOLO series algorithms have been widely used in the field of intelligent transportation because of their high precision and high speed, such as license plate recognition [15]. The latest version of YOLOv5 has better performance than all previous versions. Therefore, we take YOLOv5 as the basic model of motorcycle detection, and by modifying the depth and width of the backbone network, modifying the output of the network, integrating triplet attention, improving NMS, using K-means++ to recalculate the anchor size, we call the model used in this stage YOLOv5-MD.

The dataset for this stage was obtained from the traffic surveillance video. In the traffic scene, there are cars, motorcycles, bicycles, tricycles and other types of vehicles. We find that bicycles, forward-looking tricycles and

motorcycles are similar in their riding state. In [32] and [33], they only consider one motorcycle category, which will cause a lot of false detections. Therefore, in order to reduce the false detection rate, we detect three categories at this stage, that is, motorcycle, bicycle and tricycle. At the same time, we found that the number of bicycle and tricycle images is relatively small, so we use data enhancement methods, such as flipping, translating, blurring, etc., to expand a small number of categories.

The data of this stage is from the original image of traffic monitoring, which has a high resolution, and also has the following challenges:

The size of motorcycles at different distances from the camera varies greatly.

The images taken in different scenes have different viewing angles, such as profile, front, back and so on.

The illumination of images collected at different time intervals is different.

The challenges of different weather, such as rain and snow.

The challenges of crowded scenes, such as motorcycles blocking each other and other vehicles blocking each other.

In order to overcome the above challenges, first of all, we use a larger model, the depth_multiple and width_multiple parameters are set to 0.67 and 0.75 respectively, and triplet attention is added at the end of the backbone network to extract features of motorcycles to the greatest extent. Then, considering the real-time performance, we choose 640 × 640 network input size and adopt a multi-scale training strategy. Finally, the filter number of the last convolution layer is modified to match the number of categories. YOLOv5 uses the same output as YOLOv3, with three branches, and each branch matches three anchor boxes. Each anchor box uses four coordinates (x, y, w, h), confidence and C category probabilities. Therefore, the number of filters per branch

The detection category in this stage is three, so the convolution kernel number of the final convolution layer is 24 ((3 + 5) × 3).

Since motorcycle detection is similar to other object detection tasks, we do not start training from scratch. We used the pre-trained model obtained from the COCO dataset to fine tune the YOLOv5-MD model. In the training process, when the overlap ratio IOU between the prediction box and ground truth box is more than 0.5, it is regarded as a positive sample, otherwise, it is a negative sample. In the test phase, we choose different confidence thresholds and IOU thresholds for soft-NMS to test, and finally select the threshold that takes into account the recall rate and accuracy rate as our best result.

CONCLUSION

In this paper, we introduce a real-time end-to-end helmet detection of motorcyclists method based on YOLOv5 algorithm. This method can automatically detect the motorcycle in the video or image, and judge whether the rider on the motorcycle is wearing a helmet. Our method includes two stages ofmotorcycle

detection and helmet detection, and for each stage, we train a model, which are

YOLOv5-MD and YOLOv5-HD, to achieve a real-time effect while ensuring high accuracy. In addition, our model size is still very advantageous, the first stage model is only 20.6 MB, and the second stage model is only 14.8 MB. In order to verify the effectiveness of our method, we also propose a new motorcycle helmet dataset, HFUT-MH, which is obtained from multiple traffic scenes in China with a variety of complex weather conditions, different occlusion conditions and different light conditions. In our dataset, the final end-to-end motorcycle helmet detection mAP reached 97.7%, F1-score reached 92.7, detection speed reached 63 FPS, achieved high accuracy and real-time effect. In addition, our method can determine whether the motorcycle is overloaded by calculating the number of helmets and No_helmets. In the future, we may add a tracking algorithm to it, and detect the same object only once to avoid repeated detection.

VI. REFERENCES

1. Chen, Yi-Hsiang, et al. "Pedestrian detection: A survey." IEEE Transactions on Pattern Analysis and Machine Intelligence 40.4 (2018): 972-990.

- This survey paper provides insights into object detection technologies, which can be relevant to helmet detection systems.

2. Ranjith, B. S., and M. Rajasekaran. "A Review on Recent Developments in Helmet Detection Systems for Safety of Motorbike Riders." Proceedings of the International Conference on Inventive Computing and Informatics (ICICI 2020). Springer, Singapore, 2021. 237-246.

- This paper specifically focuses on helmet detection systems, discussing recent developments and challenges.

3. Khanna, M. "Helmet Detection and Abnormal Behavior Detection System for Bikers Using Deep Learning." 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS). IEEE, 2020.

- This conference paper explores the application of deep learning techniques for helmet detection and related safety systems.

4. World Health Organization (WHO). "Road traffic injuries." Fact Sheet. Updated August 2021.

- The WHO provides statistics and information on road traffic injuries, including the importance of helmet use for reducing head injuries among motorcyclists.

5. National Highway Traffic Safety Administration (NHTSA). "Traffic Safety Facts: Motorcycle Helmets." DOT HS 812 492. November 2020.

- NHTSA provides data and research on motorcycle helmet use and its impact on reducing fatalities and injuries in motorcycle crashes.

6. Liu, Xingang, et al. "SSD: Single Shot MultiBox Detector." European conference on computer vision. Springer, Cham, 2016.

- This paper introduces the Single Shot MultiBox Detector (SSD), a popular object detection algorithm that could be relevant to helmet detection systems.